MAY 2 8 1982

504

HE 20.6208:St2/2

Draft Guidelines for Statistics and Information on Effects of the Environment on Health

ORIGINAL

COMPLETED

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES • Public Health Service • Office of Health Research, Statistics, and Technology

Draft Guidelines for Statistics and Information on Effects of the Environment on Health

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service
Office of Health Research, Statistics, and Technology
National Center for Health Statistics

Hyattsville, Maryland
December 1980

Blank Page

FOREWORD

The legislative authority for developing these guidelines is contained in Sections 306 (e)(2)(A),(B) and (E) of the Health Services Research, Health Statistics and Health Care Technology Act, Public Law 95-623 (42 U.S.C. 242k):

> "(2)(A) The Secretary, acting through the Center, shall establish, not later than two years after the date of the enactment of this subsection, guidelines for the collection, compilation, analysis, publication, and distribution of statistics and information necessary for determining the effects of conditions of employment and indoor and outdoor environmental conditions on the public health. Guidelines established under this subparagraph shall not (i) authorize or require the disclosure of any matter described in section 552(b)(6) of title 5, United States Code, and (ii) authorize or require the disclosure of any statistics or other information which is exempt from disclosure pursuant to subsection (a) of section 552 of title 5, United States Code, by reason of subsection (b)(4) of such section. The guidelines shall be reviewed and, if appropriate, revised at least every three years after the date they are initially established. Guidelines shall take effect on the date of the promulgation of the regulation establishing or revising the guidelines or such later date as may be specified in the guidelines.
>
> "(B) The guidelines shall be designed--
>
> "(i) to improve coordination of environmental and health studies, statistics, and information, and to prevent overlap and unnecessary duplication with respect to such studies, statistics, and information;
>
> "(ii) to assure that such studies, statistics, and information will be available to executive departments responsible for the administration of laws relating to the protection of the public health and safety or the environment;
>
> "(iii) to encourage the more effective use by executive departments of such studies, statistics, and information;
>
> "(iv) to improve the statistical validity and reliability of such studies, statistics, and information; and
>
> "(v) to assure greater responsiveness by the Department of Health, Education, and Welfare and other executive departments in meeting informational and analytical needs for determining the effects of employment and indoor and outdoor environmental conditions on public health.
>
> .
> .
> .
>
> "(E)(i) Each executive department shall comply with the substantive and procedural requirements of the guidelines.
>
> "(ii) The President shall by Executive order require each executive

> department to comply with requests, made in accordance with the guidelines, by the Secretary, the Administrator of the Environmental Protection Agency, the Consumer Product Safety Commission, or the Secretary of Labor for statistics and information. "(iii) The President may be Executive order exempt any executive department from compliance with a requirement of the guidelines respecting specific statistics or other information if the President determines that the exemption is necessary in the interest of national security."

A comprehensive set of guidelines for environmental statistics and information should deal with measures of human health, physical measures of the environment, and the linkage of the two. In accordance with the legal provision which allows review and revision at least every three years, this initial version treats only the first of the three areas. Subsequent versions will include consideration of environmental data and the interrelation of environment and health measures as contributions from experts in these areas are incorporated.

We appreciate the assistance of the Subcommittee on Environmental Health Statistics of the National Committee on Vital and Health Statistics in the formulation of these guidelines, as well as in reviewing preparatory versions. In addition, selected individuals from the following governmental organizations reviewed preparatory drafts: the Health Care Financing Administration, Department of Commerce, Environmental Protection Agency, Department of Labor, Consumer Product Safety Commission, Council on Environmental Quality, National Institute of Occupational Safety and Health, National Heart Lung and Blood Institute, National Institute of Environmental Health Sciences, and the National Cancer Institute.

Comments on these guidelines have been solicited from the general public through a notice in the Federal Register. Submitted comments will be used

in preparing the final version of the guidelines. In addition, these "DRAFT Guidelines for Environmental Health Statistics and Information" will be recirculated amongst various federal agencies for additional recommendations. Comments may be sent to Mr. Ronald W. Wilson, Director, Division of Environmental Epidemiology, National Center for Health Statistics, Office of Health Research, Statistics and Technology, Office of the Assistant Secretary for Health, DHHS, 3700 East-West Highway, Hyattsville, Maryland 20782.

Blank Page

CONTENTS

FOREWORD..... i

BACKGROUND..... 1

GUIDELINES FOR COLLECTION

Introduction..... 6
Guidelines..... 7
Study Plan..... 7
Recommended Data Items..... 8
Other Data Items..... 11
Sample Survey..... 13
Interviews and Questionnaires..... 14

GUIDELINES FOR COMPILATION

Introduction..... 16
Guidelines..... 16
Coding..... 16
Editing..... 18
Quality Control..... 18
Nonresponse..... 19

GUIDELINES FOR ANALYSIS

Introduction..... 21
Guidelines..... 22
Appropriateness..... 22
Errors..... 22
Analytic Methods and Statistical Procedures..... 23
Computer Software..... 25
Interpretation..... 25

GUIDELINES FOR PUBLICATION

Introduction..... 26
Guidelines..... 27
Statistical Design..... 27
Measurement Process..... 28
Quality of Data..... 30
Technical Review..... 31
Timeliness..... 31

GUIDELINES FOR DISTRIBUTION

Introduction..... 32
Guidelines..... 32
Raw Data..... 32
Publications..... 33

REFERENCES..... 35

APPENDIX A

Drafting the Questionnaire..... 38

APPENDIX B

The Pretest..... 40

APPENDIX C

Privacy Act Limitations on Accessibility of Data Records..... 42
Internal Revenue Service (IRS)..... 43
Social Security Administration (SSA)..... 43
Health Care Financing Administration (HCFA)..... 44
Office of Personnel Management (OPM)..... 44
Veterans Administration (VA)..... 44
National Center for Health Statistics (NCHS)..... 45
Vital Records..... 46
Employee Records..... 47
Disease Registries..... 49
Military Records..... 49
Other Medical Records..... 50

GLOSSARY..... 51

DRAFT GUIDELINES FOR STATISTICS AND INFORMATION ON EFFECTS OF THE ENVIRONMENT ON HEALTH

BACKGROUND

Heightened public awareness of environmental effects on health gave rise to a number of acts of Congress during the past decade. This legislation included the Health Services Research, Health Statistics, and Medical Libraries Act of 1974 (Pub. L. 93-353),[1] which put into law what was already in fact. It authorized the National Center for Health Statistics (NCHS), part of the Public Health Service, to collect statistics on the extent and nature of illness and disability in the population of the United States; the impact of health on the economy; determinants of health; environmental, social, and other health hazards; and aspects of health care and resources, utilization, and costs.

The scope provided by this legislation was expanded by the Health Services Research, Health Statistics, and Health Care Technology Act of 1978 (Pub. L. 95-623)[2] to include specific responsibilities concerning environmental health data, statistics, and epidemiology. In part, Pub. L. 95-623 was intended to address a number of problems resulting from earlier environmental legislation.

One of the most obvious concerns was that health standards promulgated under previous legislation were often challenged, primarily because of the inadequacy of scientific and medical evidence of the relationship between environmental conditions and presumed deleterious health effects. In addition, there were problems with inability to integrate meaningfully the myriad of

information items already collected. The lack of a plan to forge ahead with collection processes that would adequately satisfy existing gaps and deficiencies was sorely felt.

Other areas of concern, such as the need to improve analytic capability, were discussed in the committee report[3] that accompanied Pub. L. 95-623.The hearings also focused on problems of interagency relations within the Federal Government. Particular issues include the need to ensure availability of data to any Federal agency charged by Congress with the duty to protect public or employee health, the need for responsiveness in terms of eliminating needless "bureaucratic barriers," and the desirability of enhancing agencies' interactions as to the linkage of their data bases.

The general concern of the legislature is embodied in the testimony of Rep. George E. Brown, Jr., before the House Subcommittee on Health and Environment. He presented his findings as Chairman of the House Science and Technology Subcommittee on the Environment and the Atmosphere:

> "...During the past session, my subcommittee made a review and evaluation of environmental research programs in several agencies. Among our several findings: First, in order for us to measure the effectiveness of our environmental programs and to guide research management decisions, it is essential that we have access to better health statistics and epidemiological information.... A number of studies, reports and hearings before congress...show that our knowledge about the relationship between environmental quality and health is too incomplete to judge adequately the health effects from environmental stress. It is clear that there are adverse health impacts, but they cannot be quantified accurately. Almost unanimous agreement exists on the need for better information from environmental monitoring and more epidemiological data on pollutant caused diseases. Without such information and data, we cannot accurately evaluate our environmental quality, implement sound policy decisions, or predict future environmental trends."[4]

A number of mandates in the area of environmental health were specified by Pub. L. 95-623.[5] Among these are the following:

- The development of a plan for collecting and coordinating statistical and epidemiologic data on the effects of the environment on health. This mandate has been fulfilled, and the plan has been published in a report[6] that includes a review of environmental health data collection systems, an assessment of data needs for environmental epidemiology, and recommendations for dealing with deficiencies in environmental health data collection and for coordinating Federal environmental health activities.

- A joint effort between the National Center for Health Statistics and the Institute of Medicine of the National Academy of Sciences to study the costs of environmentally related diseases and adverse effects on humans. A plan for an ongoing study is currently being prepared that will consider problems such as (a) identification of the pollution and the pollutants, (b) identification of associated diseases and adverse effects on humans, (c) identification of the source of such pollutants on a geographic basis, and (d) quantification of present and projected health costs of the diseases and effects.

- A study of the issues in establishing a Federal system to assist in locating individuals who may have been exposed to hazardous substances, determining the effect of such exposure on their health, and providing them access to appropriate medical care and treatment.

These guidelines for the collection, compilation, analysis, publication, and distribution are divided into guidelines for data and statistics dealing with measures of human health, physical measures of the environment, and the linkage of the two. In accordance with the section of Pub. L. 95-623

that provides for review and revision at least every 3 years, this initial version treats only the first of the three areas. Subsequent versions will include consideration of environmental data and the interrelationship of environment and health measures as contributions from experts in these areas are incorporated.

The guidelines in this report deal only with observational data on human health, for example, those that are obtained in ecologic, cross-sectional, case-control, and cohort studies.* Data from experimental animal and clinical studies are considered beyond the scope of this effort and have been treated elsewhere.[7,8,9] Since the activities of the National Center for Health Statistics traditionally have centered around population surveys, considerable emphasis is placed on topics specifically related to population surveys rather than to other types of studies, such as case-control or cohort studies. Publication of these initial guidelines is expected to elicit comments and contributions from experts in the latter areas, which can then be incorporated in subsequent versions. These guidelines are not meant to serve as a comprehensive manual on conducting observational studies.

For compliance with Pub. L. 95-623, the format of the guidelines will follow those five topics specifically mentioned in the authorizing provision—collection, compilation, analysis, publication, and distribution. However, the importance and amount of effort involved in the planning and design phase that precedes the five specified phases cannot be overlooked. This topic is treated in the section on collection.

---

*See Glossary for definitions of technical terms.

Numerous Federal standards and guidelines exist to address particular aspects of environmental and health statistics, notably the directives of the Office of Federal Statistical Policy and Standards, Department of Commerce,[10] which are the basis for these guidelines. That office has the legislative authority to perform the tasks outlined in Section 306 (2) of the Public Health Service Act, and it has issued guidelines prescribed by directives that are to be followed by all executive departments engaged in the direct collection, compilation, and publication of statistical data.

GUIDELINES FOR COLLECTION

Introduction

This section deals with general activities associated with collecting data on human subjects in population surveys and in epidemiologic studies.* The topics that are included are not meant to be exhaustive; they are suggestive of the wide range of activities in the area of collection.

First, the development of a study plan and design is discussed. This is a necessary, initial step preceding data collection. Much has been published on this phase, and considerable time spent on training statisticians and epidemiologists in this area. Its inclusion here serves not to repeat nor to encompass but to acknowledge the importance of a plan and design as a preliminary to data collection activities.

In the remainder of this section, guidelines for the collection of data are presented. These guidelines suggest methods to (1) promote comparability of information so that relevant subcategories, (that is, age, race, and sex) are available in all data collections, thus enabling direct comparisons of statistical profiles grouped into subcategories; (2) expand potential linkage opportunities among different data sets, using age, race, and sex together with a common identifier, such as social security number; (3) increase the availability of data on major confounding factors such as smoking with respect to personal health outcome; and (4) extend the ability to determine environmental exposures through the collection of residential and occupational histories.

---

*See Glossary for definitions of technical terms.

Guidelines

1. Study Plan

Data collection should be preceded by a statement describing the purpose for the study and a plan for analysis. This statement should include the following:

a. purpose, which is best established through collaboration between subject specialists and statisticians or epidemiologists;

b. potential significance and utility of the proposed work, in particular, the effect upon the acceptability and uniformity of data available to public health planners and researchers;

c. implications for health-related Federal policy, if policy-related decisions are to be affected;

d. actual work to be performed, including kinds of data to be obtained and anticipated statistical analyses to be performed;

e. limitations and pitfalls of the proposed procedures;

f. design, including whether the study is cross-sectional, panel or longitudinal, cohort, etc.; whether the study is to be descriptive—e.g., survey—or analytic; and if analytic, whether it is to explore relationships or attempt to establish etiology;

g. level of precision necessary to achieve meaningful results giving special consideration to sample size;

h. methods and techniques to be employed for data collection, providing specifications as to whether personal interview, telephone interview, or mail questionnaire will be employed; whether the sampling units will be households or individuals; rules as to respondent eligibility; criteria for review of records, etc.;

i. cost estimate broken down into study stages, including recruitment, data collection, analysis, publication, distribution, etc.;

j. time frame for completion of various stages, including governmental clearance and publication.

2. Recommended Data Items

These items are recommended to ensure that data are collected on a minimum set of variables often associated with health status, to improve the comparability between studies of data collected on the minimum set of variables, and ultimately to improve the comparability of risk statistics derived from epidemiologic studies. For the specialized purposes of any given study, more detailed categorization can be used. However, these detailed categories should be capable of being collapsed into the following recommended data items:

a. Date of birth should be recorded as "Month, Day, Year" and should be filed in accordance with the Federal Information Processing

Standards (FIPS) on calendar date recording,[11] with at least two digits recorded for each element of the date.

b. Sex should be recorded as either "Male" or "Female."

c. The standard for race and ethnicity that has been established by the Office of Federal Statistical Policy and Standards (Directive No. 15)[10] is recommended for collection of this information. Information on race and ethnicity should be collected separately in accordance with the first alternative of Directive No. 15, which states: "To provide flexibility, it is preferable to collect data on race and ethnicity separately. If separate race and ethnic categories are used the minimum designations are: (1) Race: American Indian or Alaskan Native, Asian or Pacific Islander, Black, and White. (2) Ethnicity: 'Hispanic origin' or 'Not of Hispanic origin.' "

The basic definitions for these categories are as follows:

American Indian or Alaskan Native.—A person having origins in any of the original peoples of North America who maintains cultural identification through tribal affiliation or community recognition.

Asian or Pacific Islander.—A person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Philippine Islands, and Samoa.

Black.—A person having origins in any of the black racial groups of Africa.

Hispanic.—A person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish culture or origin, regardless of race.

White.—A person having origins in any of the original peoples of Europe, North Africa, or the Middle East.

d. The effect of cigarette smoking on personal health is well established. It is difficult to imagine epidemiologic studies involving interview of individual respondents in which this information should not be collected. Ideally, a smoking history for an individual should be a complete, chronological history. However, such an approach often is not feasible. Therefore, it is recommended that if smoking information is obtained at all at least the following elements should be noted:

Ever smoked.—A smoker is a person who has smoked 100 or more cigarettes in his or her entire life.

Age first started smoking.—Age in years at the time the person first smoked regularly.

Current smoker.—Does the person smoke now?

Current amount smoked.—For current smokers, this is the average number of cigarettes smoked per day.

Heaviest amount ever smoked.—For current and former smokers, this is the average number of cigarettes smoked per day during the time they smoked most heavily.

Age quit smoking.—For former smokers, this is the age in years at which they last smoked.

Information as to the type of cigarettes smoked is desirable, but standardization of categorization has not been achieved.

e. Social security number or some single identifying number for individuals should be recorded. While it is recognized that the social security number is not necessarily unique and that not every individual in this country has such a number, it is commonly available, known, and used for identification purposes. Therefore, lacking a better identifier, its use is recommended at this time. In conjunction with the name and date of birth of an individual included in a study, the social security number will permit access to the National Death Index. The National Death Index is a mechanism operated within the National Center for Health Statistics for the express purpose of helping researchers determine whether a subject in their research study died and, if so, which State can be contacted to obtain a copy of the death certificate and to learn about the circumstances of the death. The collection of identifying information should be in accordance with Privacy Act[12] limitations.

3. Other Data Items

In general health surveys of human populations, investigators should consider collecting the following information to determine possible environmental exposures:

a. Personal health histories are particularly valuable, as for example in occupational studies. However, in collecting health histories as opposed to collecting data on current health status, allowances

should be made for the fact that respondents often have difficulty recalling occurrences as well as times of acute and chronic episodes. If information is obtained over a short-recall period, care must be given to the extent to which one can infer a long-term history. To stimulate recall response uniformly, it is recommended that medical conditions be selected from a predetermined checklist.

If prospective studies are planned, consideration should be given to using the NCHS National Death Index to facilitate tracking the mortality experience of panel cases. This procedure requires collecting the social security number to ensure an adequate match with future death records.

b. A detailed, chronological occupational history that would include name of employer, job title and job description, dates, address of work site for geocoding, as well as identification of toxic substance to which presumably exposed, would constitute the ideal set of occupational data. As yet, methods have not been developed and refined to overcome the difficulties in acquiring such a data set. However, the following procedures are recommended to enhance the prospect. A checklist of specific hazardous occupations and substances should be developed to fit the specific purposes of the study. Particular attention should be given to the size of the study population to determine if there would be enough cases in any given exposure category to be statistically meaningful. Questions should also be considered that would relate to personal knowledge of toxic exposure and to job changes resulting from hazardous health conditions.

c. Residential histories can be an integral part of occupational studies. As with occupational histories, a complete residential history is ideal but difficult to collect, code, and analyze. Elements essential for residential codes include name of place, inside or outside city limits, county, and State. Consideration should be given to the use of ZIP codes, which has obvious limitations in any but short-term recall studies. "Place of birth" should be the residence of the mother, not the location of the birth.

4. Sample Survey

a. The investigator should follow accepted practices of sample survey theory and methodology as given in texts like Hansen, Hurwitz, and Madow,[13] Kish,[14] Cochran,[15] and Cassel, Sardnal, and Wretman.[16] It is important that the investigator have an adequate sample size and that the design be appropriate to the research plan. Almost every study will dictate its own sample survey features. The survey design and features, including the methodology for calculating sample size, should be documented and maintained for public access.

b. The population of interest or study population should, where possible, coincide with the sampled population. Since it is often difficult to sample from the population of interest, investigators may sample a population that is slightly different but related. The conclusions drawn from the study then apply to the sampled population, and caution must be used in associating these conclusions with the population of interest. The sampling frame, or list of eligible individuals or elements from which the sample is selected, should be specified in the documentation.

c. When sampling from a population, all units in the population must have a chance of being sampled and observed in order to provide a representative sample. The chance of selection does not have to be the same for every unit in the population, but each must have a known non-zero chance of being selected into the probability sample. Exclusion of any individual or group from availability for selection yields a nonrepresentative sample that is different from the population of interest.

5. Interviews and Questionnaires

a. In federally funded studies where interviews are to be conducted to obtain data, advance contact should be made before the interview to inform the subjet of some basic information, such as the legislative authorization of the study, the purpose of the study, the type of information to be solicited, and the voluntary or mandatory nature of participation in the study. Conformance with the Privacy Act [12] should be checked.

b. Appendix A lists basic principles that should be considered when drafting the questionnaire. Concepts forming the background for current practices or Federal statistical agencies and measurement methods advocated are described in the Household Survey Manual[17] and in Basic Background Items for U.S. Household Surveys.[18] These publications include appropriate questions for measuring important basic characteristics of the population and furnish advice on survey operation, including criteria for obtaining reliable responses.

c. Before "going into the field" to collect data with a questionnaire, a pretest should be conducted to help detect previously unperceived difficulties with this phase of data collection. All associated aspects should be tested, not only the data collection instrument but administrative procedures, encoding procedures, data manipulation and processing, and so on. Appendix B lists four basic considerations for a pretest.

GUIDELINES FOR COMPILATION

Introduction

The word "compilation" is defined here as those steps taken to translate data collected in a study to a form usable by the investigator. Typically, this form is a machine readable, magnetic computer tape. The steps necessary are coding the data items contained in the collection instrument, entering the data into the medium of storage, editing the data set, and imputing for missing observations.

Guidelines

1. Coding

Data collected in the information gathering process are often cast into another form for a variety of reasons, such as for tabulation, analysis, or storage. This involves transposing raw data to another form, usually according to a given classification scheme (which involves to some degree an abstraction process) or to a coding scheme (usually a translation of words to numbers).

Although slightly different, the use of classification and coding schemes will be treated together here in the sense that they are meant to be representations of the original data.

a. Widely used disease coding schemes are the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) and the Z16.2 codes of the American National Standards Institute (ANSI). However, for morbidity and other types of health data such as symptoms,

signs, disabilities, and test results, other systems exist and are used. The system used, complete with version number, should be identified adequately so that code and condition can be translated unambiguously by independent users.

For better coordination of existing and future data collection systems, documentation of classification scheme should include:

(1) complete identification of the health data items being classified or coded and the classification or coding system being used;
(2) compilation or collection of special instructions and coding manuals;
(3) identification of available conversions or translations to other classification schemes and a statement of comparability to other data sets wherever possible.

b. To promote uniformity in the coding of geographic areas, it is recommended that the codes in the National Bureau of Standards publications be used. Federal Information Processing Standards (FIPS) Publication 5-1[19] contains codes for States and outlying areas and Publication 6-2[20] contains codes for counties and county equivalents of the States of the United States.

c. To promote uniformity in the coding of occupational history data, it is recommended that the Standard Occupational Classification Manual (SOC)[21] be used for occupational classification and the Standard Industrial Classification Manual (SIC)[22] be used for industry classification.

2. Editing

Editing data reported from the survey has two general purposes:

- to minimize the differences of reporting on individual questionnaires to the extent that care and reasonable judgments based on the evidence at hand make possible, thereby collecting data that are as accurate as circumstances permit;
- to maintain running tabulations of numbers of errors discovered, classified by type and by cause, so that the investigator can be guided in reducing mistakes in later cycles of data collection.

The mechanics for accomplishing these purposes will vary with the type of data gathering instrument and data processing equipment. Another factor is the availability of collateral data useful for checking the acceptable magnitude or range of entries for particular items.

To edit the data set, the investigator, insofar as possible, should decide in advance how to handle faulty data, such as omissions, incomplete entries, and incorrect or inconsistent information. Procedures used to edit the data set should become part of the study documentation.

3. Quality Control

It is advisable to establish standards of quality in coding and editing. The investigator should measure the quality of the coding and editing operations and take appropriate action if they are not at an acceptable level. Procedures used in the quality control program should become part of the study documentation.

4. Nonresponse

Nonresponse is typically classified as unit nonresponse or item nonresponse. Unit nonresponse includes eligible sample respondents who were never located, those who died, those too infirm to be interviewed, those who refused to participate, and those for whom measurements were obtained but were lost before processing. Item nonresponse includes eligible sample respondents for whom incomplete measurements were obtained. Incompleteness may result from interviewer error (failure to ask a question), respondent omission (intentional or unintentional failure to answer a question), and so forth.

Virtually every survey will contain a certain amount of unit nonresponse and item nonresponse. Available respondents should be recontacted to obtain the needed information; not doing so would result in elimination or adjustment, neither of which are likely to yield the same statistical result as an actual response. For those who are unavailable because of death or illness, consideration should be given to criteria for admissible proxy response. If no adjustment for nonresponse is made, this implies that the nonrespondents are like the respondents with regard to the measurements of interest. If the two groups are different, however, failure to adjust introduces bias* proportional to the degree to which respondents and nonrespondents differ.

Upon completion of the data collection stage, the degree of nonresponse should be quantified and the possible causes considered. Provided the nonresponse does not indicate a flaw in the basic structure of the study, attempts should be made to ameliorate the situation. These steps include recontact of the

*See Glossary for definitions of technical terms.

respondents, elimination of the unit from study, or statistical estimations of the missing data if item nonresponse is the case.

In population survey samples, adjustment for nonresponse is usually made by imputing[23] measured values of respondents to the nonrespondents. The assumption that nonrespondents are like respondents is reasonable if it can be shown that the groups are similar in characteristics that are correlated with measurements being obtained in the survey. Procedures used to impute for missing values should become part of the study documentation.

GUIDELINES FOR ANALYSIS

Introduction

In this section as in others, several issues exist. Descriptive studies can differ fundamentally from investigative analyses; even the language used to treat them can mirror these differences. For example, investigative analyses are usually "interpreted," while descriptive statistics are "observed" or "discussed."

Even more importantly, analysis is not simply a computational process. The selection of a statistic and the selection of a test for that particular statistic are based on consideration of many components of the study, for example, the sampling procedure, the method of data collection, the problems encountered in the data collection process, the resulting data deficiencies, and the subject under study. The analysis itself should be a factor in determining these other facets of the study. No two studies are alike in all these components. For this reason and to allow for versatility and creativity in the analysis, the material presented here is intended for use as a general guide in performing analyses.

This section recommends that: (1) an analysis be appropriate, (2) potential sources of error and bias be made explicit where possible, and (3) results be interpreted or described adequately and accurately. The language is intentionally nontechnical to accommodate more general usage.

Guidelines

1. Appropriateness

Statistics chosen and analyses performed should be appropriate to the purpose of the study. For example, descriptive analyses usually include a measure of central tendency and a measure of scatter, while investigative analyses might include statistics based upon underlying models and basic assumptions. Occasionally, in statistical problems with unique properties, the completion of the analysis depends upon further developmental work in statistical methodology. In each case, documentation should include statements that indicate the reasons for the statistics chosen and the analysis performed.

2. Errors

Errors can be considered as limitations on resulting statistical estimates. One major goal of a comprehensive study design is to eliminate avoidable errors, for example, failure to obtain a representative sample resulting in bias, as decribed in "Guidelines for Collection" under 4. Sample Survey, b. However, there are other sources of error that are inherent in statistical procedures and are therefore unavoidable. These should be documented and quantitatively estimated. Examples of such procedures include reporting of sensitivity* and specificity* levels of classification, subject response variation, and intraobserver and interobserver variation. For a detailed technical discussion of errors, reference should be made to the set of standards[24] published by the U.S. Department of Commerce. These standards include methods of presenting sampling and nonsampling errors.

*See Glossary for definitions of technical terms.

Computational errors also should be avoided. With the advent of computers, estimates are available for errors that occur in rounding,* which are a function of the computational algorithm used in the computer program or the bit-precision of the computer. Estimates should be provided for such errors, and statements should be included as to the effect upon the calculations, statistics, and ultimate decisions.

3. Analytic Methods and Statistical Procedures

The recommendations that follow are taken in part from the documentation guidelines for epidemiologic studies[25] being prepared by the Interagency Regulatory Liaison Group.

A description should be given of the procedures and analytic methods used for estimation and for testing specific hypotheses. This description should include:

a. a statement of the underlying assumptions of the procedures and analytic methods used;

b. an assessment of possible bias (direction and magnitude) in the analytic methods used;

c. criteria for disease classification and choice of disease groupings. If different procedures are used to classify disease in the study and comparison subjects, the effects of the procedures and an assessment of any bias in the study results should be given;

*See Glossary for definitions of technical terms.

d. a description of the methods used to account for nonrespondents and attrition. A discussion should be given of the assumptions made, their justification, and the direction and magnitude of anticipated bias;*

e. an evaluation of the effect of potential confounding factors (e.g., age, sex, ethnic group, lifestyle). In particular, a description should be given of any methods used to take account of these confounding factors (e.g., adjustment or matching);

f. a designation of the statistical tests and/or the method of estimation used in the study, including the assumptions underlying the test or estimation procedures and the appropriateness of the assumptions. Specific formulas should be given, where appropriate, and literature references presented;

g. justification for combining subgroups of study or comparison subjects when multiple sources of data are used (e.g., data from multiple investigations or multiple populations). The degree to which the different sources differ in definition, data collection procedures, time frames, estimating procedures, and sampling frames should be specified. Where differences exist, efforts should be made to determine the effects on the statistical results and explanations should be attempted.

---

*See Glossary for definitions of technical terms.

4. Computer Software

If a computer software package is used for analysis, it should be identified. The accuracy of computer software programs used for analyzing the study data set should be verified by using a test data set.

5. Interpretation

Results from an analysis should be interpreted as completely as possible in relation to the stated objectives of the study, and if appropriate, comparisons should be made with results from other studies. Special consideration should be given to limitations of the data set and the specific methods of analysis used for the study. Interpretation of results and inferences drawn should be supported by a discussion of probable bias, biological plausibility, and consistency in relation to other studies.

For example, an investigative analysis in which determination of a cause and effect relationship is attempted should provide results that form the basis of a discussion and evaluation of the following kinds of issues: strength of the association, sequence of cause and effect events, latency, consistency of findings over time and among different populations, and biological plausibility of the relationship.

GUIDELINES FOR PUBLICATION

Introduction

The term "publication" refers here to those media through which information relating to environmental health statistics is made known. Ordinarily, this refers to printed materials such as books, magazines, and periodicals. There is, however, great latitude of meaning of the term, especially with the application of computer technology in the field of communications.

The recommendations in these guidelines are for materials to be included in publications intended to be complete in themselves or self-contained. Obviously, applicability would be affected if the publication is intended for a special purpose, such as, a response to a specific question or the production of bibliographies. However, the inclusion of guidelines for special purpose publications is considered beyond the scope of this work.

For greater detail in specific areas, reference should be made to recommendations in the Manual of Standards and Procedures for Reviewing Statistical Reports[26] and to the Manual of Statistical Presentation.[27] Guidelines for documentation of epidemiologic studies are also given elsewhere.[25] Every analytic report should contain a detailed description of the data set, either in the report itself or by reference to an existing, available publication that contains complete details. This description should treat the statistical design for sampling, the measurement process, and data reduction, estimation, and quality.

Guidelines

1. Statistical Design

Every study should have a document that gives a complete description of the study's activities including, for example:

a. purpose of the study;

b. target population, describing the population to which statistical estimates pertain;

c. source of data, giving the name of the organization that collected the data set and a general outline of the method of data collection;

d. sampling frame, listing the totality of units from which the sample is taken;

e. time reference, including the time period during which the data were collected as well as the reference period for the data, such as the number of weeks preceding the interview and the time worked in a particular job;

f. comparability, between data dealt with in the publication and those collected or reported elsewhere, in particular, related large-scale studies or other major sources of environmental data or significant prior studies;

g. sampling units, Primary Sampling Units (PSU's), segments, etc.;

h. stratification and clustering, outlining procedures used to form strata or clusters and giving the number so formed;

i. sample selection process, providing procedural descriptors, such as systematic, simple random, probability proportional to size;

j. sample size, indicating the number of units selected at each stage of selection, for example, the number of PSU's, segments, households, and persons;

k. other features that may be important to the statistical design, such as controlled selection, double sampling, and oversampling;

l. sources of potential bias;

m. utility, assessing usefulness for Federal executive departments and for other researchers and environmentalists.

2. Measurement Process

Details on the measurement process should also be available either in the publication itself or in references to previous ones. The description should include these items:

a. collection process, giving a general outline of the methods used to collect data, such as household interviews or physical examinations;

b. data collection forms (questionnaires), displaying or making available those areas of data collection that deal with information treated in detail in the report. It is not necessary that the entire form be included in the report;

c. field quality control procedures, outlining those procedures used to ensure that data collected are as reliable as possible, including training of interviewers, pretesting of data collection forms--particularly questionnaires--and followback efforts to reduce nonresponse or inconsistencies in data;

d. estimation, including applications of procedures such as weighting, poststratification, and nonresponse adjustment. If the procedure followed is generally accepted and known, actual formulas need not be presented;

e. data processing, outlining methods used for translations of information onto computer tape (for example, coding procedures, tabulation procedures, and sensing devices) or alluding to standard procedures, if applicable;

f. quality control, indicating the procedures used to control or reduce errors in coding and punching;

g. analytical methods, citing all methods used to analyze the data.

3. Quality of Data

Quality of data, at least in the form of errors that may be present, must be given. Potential errors may result from the sampling procedure itself (sampling errors) or other factors (nonsampling errors). Under sampling and nonsampling errors, the report should include:

a. definitions of terms used in the report, such as relative standard error and sampling variability;

b. methods of estimating sampling errors;

c. presentation of errors, in whatever numerical form is appropriate, indicating the effect on the estimates presented in the statistical tables and in the text;

d. nonresponse, discussing which items had high nonresponse and possible resulting bias;

e. methods of imputation, or substituting for missing data, and the possible effect on the quality of the data. Proportion of responses imputed should also be given;

f. response bias and measurement error, discussing the possible effects on statistical estimates given in the report. Specific

information on magnitude and direction of biases should be given.

4. Technical Review

In general, all technical terms should be defined clearly in the text or in the appendix. Before publication, the report should be reviewed by experts in the subject matter as well as in statistical methodology. Standards for the publication of statistics by Federal Government agencies are presented in the Statistical Policy Handbook.[10] For more technical items included in the report, such as tables and graphs, reference should be made to the specifications given in the U.S. Government Printing Office Style Manual.[28]

The review process should also be sensitive to the protection of the privacy of individual respondents, as mandated by the Privacy Act.[12]

5. Timeliness

When policymaking decisions of executive departments are to be affected, it is important that publication be timely so that the maximum potential value of the effort can be realized. Prompt release and distribution are topics dealt with in Directive No. 4,[10] "Prompt Compilation and Release of Statistical Information," of the Statistical Policy Handbook.

GUIDELINES FOR DISTRIBUTION

Introduction

As discussed here, the term "distribution" refers to sharing environmental health data and resulting information (publications) between and among Government agencies, Federal executives, health administrators, university professors, clinical researchers, and others. The concept involves both raw data and publications. It is accompanied by the assumption that the material to be distributed is scientifically valid and of a form that is easily understandable. These issues of quality and documentation have been dealt with in previous sections, so that the main issue for additional discussion in this section is that of accessibility and its corollary privacy and confidentiality.

Some of the major Federal data collection systems that contain environmental health or related data and that are accessible through tape, microfiche, or publication are listed in Environmental Health: A Plan for Collecting and Coordinating Statistical and Epidemiologic Data.[6]

Guidelines

1. Raw Data

After prompt encoding and editing, environmental health data collected by Federal agencies should be made available for public use to the extent provided by the collecting agencies' authorizing legislation.

Individually identifiable records should be made available whenever warranted by the environmental health issue under question, for example, in linking health records of individuals with data on exposure to environmental hazards or in retrospectively identifying and locating individuals exposed to suspected hazardous environmental conditions.

For use in epidemiologic research, Appendix C indicates the accessibility of large centralized data systems of the Internal Revenue Service, Social Security Administration, Health Care Financing Administration, Office of Personnel Management, National Center for Health Statistics, Veterans Administration, Department of Energy, National Research Council, Occupational Safety and Health Administration, and the National Institute for Occupational Safety and Health. It also includes comments on vital records, disease registries, and medical records.

2. Publications

Any publication of scientific merit should be in a medium that is regularly classified and incorporated into existing cataloging systems easily accessible to Federal executive departments and to the public. The publications themselves should also be readily available.

It is important for Federal agencies to coordinate activities so that the maximum audience is reached with minimal duplication. Data dissemination should be an integral part of the planning process. Agencies may share distribution networks, systems, and mechanisms. Interagency coordination should be promoted.

To ensure maximum availability and accessibility, Federal agencies should follow a uniform procedure whereby: (a) each publication, public use tape, or microform product is entered into and made available through a centralized distribution service, such as the National Technical Information Service, Department of Commerce, and (b) an abstract of each new publication is sent to Index Medicus, to Excerpta Medica, and to other appropriate abstracting systems to ensure that environmental reports are included in MEDLARS* and other bibliographic systems.

*Medical Literature Analysis and Retrieval System, National Library of Medicine, the National Institutes of Health, Bethesda, Md.

REFERENCES

1. Public Law 93-353, 93rd Congress, 1st Session (88 Stat 362).

2. Public Law 95-623, 95th Congress, 2nd Session (92 Stat 3443).

3. Report by the Committee on Interstate and Foreign Commerce, Health Services Research, Health Statistics, and Health Care Technology Act, House of Representatives, Report No. 95-1190, U.S. Government Printing Office, May 15, 1978.

4. Hearings before the House Subcommittee on Health and Environment, "Health Services Research and Health Statistics Amendments of 1978." 95th Congress, 2nd Session, Serial No. 95-117, Apr. 26-28, 1978, p. 396, 400-1.

5. P.L. 95-623, Section 8(a)(2)(A) and (B).

6. Environmental Health: A Plan for Collecting and Coordinating Statistical and Epidemiologic Data. DHHS Publication No. (PHS) 80-1248, June 1980, Washington, D.C.

7. Draft I.R.L.G. Guidelines for selected acute toxicity tests, Testing standards and guidelines work group: Interagency Regulatory Liaison Group, Aug. 1979.

8. FDA 77-3040. "General Considerations for the Clinical Evaluation of Drugs" (GPO 017-012-00245-5); Sept. 1977.

9. FDA 77-3041. "General Considerations for the Clinical Evaluation of Drugs in Infants and Children" (GPO 017-012-00246-3), Sept. 1977.

10. Statistical Policy Handbook, U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, 1978.

11. Calendar Date, Federal Information Processing Standards, Publication 4, November 1, 1968, United States Department of Commerce, National Bureau of Standards.

12. The Privacy Act of 1974 (5 U.S.C. 552a).

13. Hansen, M. H., Hurwitz, W. N., and Madow, W. G.: Sample Survey Methods, Volume I, Methods and Application. John Wiley and Sons, Inc., N.Y., 1956.

14. Kish, L.: Survey Sampling. John Wiley and Sons, Inc., N.Y. 1965.

15. Cochran, W. G.: Sampling Techniques, 2nd edition. John Wiley and Sons, Inc., N.Y., 1963.

16. Cassel, C. M., Sardnal, C. E., and Wretman, J. H.: Foundations of Inference in Survey Sampling. John Wiley and Sons, Inc., N.Y., 1977.

17. Household Survey Manual, 1969. Bureau of the Budget. Executive Office of the President.

18. Van Dusen, R.A., and Zill, N. (eds.): Basic Background Items for U.S. Household Surveys, Social Science Research Council, Center for Coordination of Research on Social Indicators, Washington, D.C., 1975.

19. "States and Outlying Areas of the United States," Federal Information Processing Standards Publication 5-1, June 15, 1970, U.S. Department of Commerce, National Bureau of Standards.

20. "Counties and County equivalents of the States of the United States," Federal Information Processing Standards Publication 6-2, September 15, 1973, U.S. Department of Commerce, National Bureau of Standards.

21. Standard Occupational Classification Manual, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, 1980.

22. Standard Industrial Classification Manual, Executive Office of the President, Office of Management and Budget, 1972 and the 1977 Supplement Office of Federal Statistical Policy and Standards, U.S. Department of Commerce.

23. Levy, P.S. and Lemeshow, S.: Sampling for Health Professionals. Lifetime Learning Publications, Belmont, Ca., 1980, pp. 258-271.

24. Standards for Discussion and Presentation of Errors in Data, Technical Paper 32, Bureau of the Census, U.S. Department of Commerce, 1974.

25. "Draft Documentation Guidelines for Epidemiologic Studies," Interagency Regulatory Liaison Group, Sept. 1979.

26. Sirken, M.G., Shimizu, B.I., French, D.K., and Brock, D.B.: Manual of Standards and Procedures for Reviewing Statistical Reports, National Center for Health Statistics, Mar. 1976.

27. Manual of Statistical Presentation, Number 10, January 1970 Statistical Analysis Survey Section, Statistics and Analysis Branch, Division of Research Grants, National Institutes of Health.

28. U.S. Government Printing Office Style Manual, Revised Edition, Jan. 1973.

29. Interagency Task Force on the Health Effects of Ionizing Radiation: Report of the Work Group on Records and Privacy, Office of the Secretary, U.S. Department of Health, Education, and Welfare, June 1979.

30. Interagency Task Force on the Health Effects of Ionizing Radiation: Report of the Work Group on Exposure Reduction, Office of the Secretary, U.S. Department of Health, Education, and Welfare, June 1979.

31. "Privacy Act Implementation, Guidelines and Responsibilities," Federal Register, Vol. 40, No. 132, Part III, July 9, 1975.

32. Data Systems of the National Center for Health Statistics, DHEW Publication No. (PHS) 79-1247, Jan. 1979.

33. Model State Vital Statistics Act and Model State Vital Statistics Regulations. DHEW Publication No. (PHS) 78-115. 1977 Revision.

Appendix A

DRAFTING THE QUESTIONNAIRE

Some basic principles to be followed when drafting the questionnaire are:

1. question wording must be simple and devoid of slang or jargon;

2. interviewer instructions should be printed on the questionnaire;

3. the questionnaire form should be orderly, neat, and uncluttered, especially when respondents are requested to make the entries;

4. the questionnaire should be as brief as possible, since respondents are more willing to cooperate when the number of questions asked is small;

5. space should be provided for inserting additional responses and for making notes;

6. question grouping and ordering should be carefully considered to ensure a flowing interview for both the respondent and the interviewer;

7. screening questions should be used to assist the interviewer in skipping inappropriate sections of the questionnaire;

8. coding for entry to computer media should be included on the form if automation is intended.

Appendix B

THE PRETEST

Before beginning the pretest operation, the following should be considered:

1. The objectives of the formal pretest of the questionnaire should be specified. If the pretest is to be a scaled-down version of the actual test, then the sampling frame for the formal pretest should reflect as much as possible the target population of the actual survey, that is, the demographic characteristics of the pretest sample should resemble those of the target population. If, however, certain questionnaire subsections are to be focused upon, the pretest sample might be heavily weighted for this purpose. The adequacy of the sampling frame should be tested for completeness, accuracy, and convenience.

2. The pretest operation should include testing of processing and analytic procedures. If an objective of the pretest is to determine which of several processing procedures or analytical techniques should be utilized, these factors should be considered as the questionnaire is being designed. Pretests should be designed so that information on nonresponse is obtained for analysis.

3. The procedures and instructions to be utilized during the actual survey should be tested the same way that they are intended to be implemented. If established procedures are involved, these should be tested concurrently with any new or modified method that is to be used. If alternate procedures are to be tested, each should be tested under similar conditions.

4. A manual should be prepared that contains instructions for the pretest interviewers on the procedures to be followed in administering the questions. The interviewers should be provided these manuals and trained in the procedures. Interviewers should be monitored or observed while administering the questions. During the pretest, interviewers should note any problems on questionnaires. In addition, interviewers should orally comment on problems during debriefing. This feedback is important and should be considered in writing the final version of the instruction manual.

Appendix C

PRIVACY ACT LIMITATIONS ON ACCESSIBILITY OF DATA RECORDS

For most Federal agencies the Privacy Act of 1974,[12] as amended, governs the release of identifiable information. Details regarding privacy and confidentiality as they affect various Federal agencies are given in the reports of the Interagency Task Force on the Health Effects on Ionizing Radiation.[29,30]

Release of identifiable information within an agency is permitted under the Privacy Act on a need-to-know basis. "Agency" is defined to include all components of an executive department. In the Department of Health and Human Services, for instance, the Health Care Financing Administration can release identifiable data to the National Cancer Institute on a need-to-know basis. The Privacy Act also allows data to be released outside an agency under the "routine use" provision, which requires only that the outside use be compatible with the purpose for which the data were collected. The statute, legislative history, and Office of Management and Budget guidelines[31] have furnished minimal guidance to Federal agencies in defining "compatibility" under the Privacy Act. As a result, the agencies have promulgated regulations providing widely varying interpretations of the term. In addition, Federal agencies may be governed by specific privacy laws providing more restrictive conditions on data release than those in the Privacy Act. Current limitations on the use of major data sources follow.

1. Internal Revenue Service (IRS)

Prior to the enactment of the Tax Reform Act of 1976, epidemiologists were able to obtain certain information from IRS records, such as the taxpayer's address on the last filed return, date of the last filed return, occupation, and if a return was filed in a particular year. However, the Tax Reform Act of 1976 substantially restricted the availability of information from tax returns and, with one exception for certain disclosures to the National Institute for Occupational Safety and Health, eliminates the use of return information for epidemiological research. The exception, which could be important for certain environmental studies, reads:

> "Upon written request, the Secretary (of the Treasury) may disclose the mailing address of taxpayers to officers and employees of the National Institute for Occupational Safety and Health solely for the purpose of locating individuals who are, or may have been, exposed to occupational hazards in order to determine the status of their health or to inform them of the possible need for medical care and treatment." (26 U.S.C. 6103(m)(3))

2. Social Security Administration (SSA)

Under the current interpretation of the Tax Reform Act of 1976, SSA may not disclose information obtained from employer tax information returns filed with IRS under the Federal Insurance Contributions Act, whether for research or for any other purpose. As a result, researchers cannot obtain listings of individuals who were employed by a particular employer.

However, SSA can forward a letter from a researcher to an individual in care of his or her employer suggesting that the individual contact the researcher. Response rates would probably be low. SSA's proposed

revised privacy regulations recognize epidemiologic research as being compatible with SSA purposes. Thus beneficiary addresses could be released, since they are not derived from IRS data.

3. Health Care Financing Administration (HCFA)

HCFA privacy regulations are currently being drafted, with publication planned for 1981. The Privacy Act applies to Medicare data, but Medicaid data mainly reside within the State and are subject to State laws and regulations. HCFA laws and regulations specify minimum standards for such State rules. In general, HCFA policies have been supportive of environmental health research.

4. Office of Personnel Management (OPM) (formerly the Civil Service Commission)

Researchers indicate that records have been made available for research, but OPM has not published an authorization for routine use disclosure of records for research purposes except for personnel research activities. Presumably, this is based on an OPM determination that personnel research is a compatible use.

5. Veterans Administration (VA)

VA permits, as a routine use, the disclosure of medical records containing the names and addresses of the individuals to any Federal agency if disclosure is necessary for the conduct of Government research in order to accomplish a statutory purpose of the agency. However, VA's own confidentiality statute prohibits VA from disclosing the names and addresses of present or former personnel of the armed services

or their dependents to researchers not employed by Federal agencies or by certain State public health agencies. Accordingly, while the VA will permit the disclosure of medical record data without identifiers to researchers outside the Federal Government, such researchers are not able to link these data to other information sources.

6. National Center for Health Statistics (NCHS)

NCHS operates a diverse survey and inventory program to collect statistics on the extent, nature, and impact of illness and disability in the population of the United States; on health resources, utilization, and costs; and on family formation, growth, and dissolution. The surveys are described in detail in Data Systems of the National Center for Health Statistics.[32] The data systems are based on national sample surveys to estimate the health of the Nation. As such, their utility for environmental health analyses differs considerably from the utility of data systems having a problem-oriented approach. An analysis of the utility of NCHS systems is given in Environmental Health: A Plan for Collecting and Coordinating Statistical and Epidemiologic Data.[6]

Data collection activities of NCHS are subject to the specific provisions of both the Privacy Act of 1974 (5 U.S.C. 552a) and the Health Services Research, Health Statistics, and Medical Libraries Act of 1974, which enacted in particular section 308(d) of the Public Health Services Act, 42 U.S.C. 242m(d) (P.L. 93-353). NCHS authorizing legislation prevents information obtained in the course of NCHS statistical activities from being used for any purpose other than that for which it was supplied. The legislation also specifies that such information

may not be published or released in other form if the establishment or person supplying the information or described in it is identifiable, unless such establishment or person has consented to its publication or release.

NCHS releases data in several ways. One way is through microdata tapes containing survey data (without identifiers). Tape availability and descriptions of tape contents are detailed in the annual publication Standardized Micro-Data Tape Transcriptions.

Another way is through publications, foremost of which are the detailed reports in the Vital and Health Statistics Series, which contain analyses of the various data systems, and the Advance Data reports, which contain timely but brief summaries of newly available data on topics of special interest.

The Current Listing and Topical Index to the Vital and Health Statistics Series, containing a cumulative list of these reports, is published annually.

7. Vital Records

Records of births, deaths, divorces, and marriages are an important source of information for epidemiologic researchers. These records are collected under State laws, and they remain under the ownership and custody of the State governments. Abstracted records are obtained by NCHS under contracts with the States; these contracts prescribe and delimit the uses that NCHS may make of the data. Section 308(d)

of the Public Health Service Act (42 U.S.C. 242m) limits NCHS to using data only for the purposes for which they were supplied. Furthermore, this section precludes NCHS from publishing information or releasing it in other forms "if the particular establishment or person supplying the information or described in it is identifiable unless such establishment or person has consented (as determined under regulations of the Secretary) to its publication or release in other form."

State responsiveness to researcher requests for data varies, and it is not unusual for restrictions to be placed on the use of information obtained from a death certificate. (Contact with next-of-kin is sometimes prohibited, for example.) In some States, the basis for access by researchers is unclear because of general or vital records privacy legislation. The NCHS Model State Vital Statistics Act[33] removes this ambiguity by making clear that research access is authorized if there are suitable safeguards.

8. Employee Records

Department of Energy (DOE).—DOE records and DOE contractor records are subject to the Privacy Act. Personnel, exposure, and medical records are thus available to DOE researchers on a "need-to-know" basis. They would also be available to other researchers if DOE were to publish guidelines for routine-use disclosure, since it is likely that disclosure of these records for epidemiologic research would meet even the strictest test of "compatibility." DOE is now in the process of assessing the availability of records, types of data included in and location, quality, and completeness of the records.

Nuclear Regulatory Commission (NRC).—The few exposure reports that are submitted to NRC are covered by the Privacy Act and are available to Federal and State agencies involved in monitoring or evaluating radiation exposure under an existing routine-use publication. Most NRC licensee records are not subject to the Privacy Act. In the absence of specific regulatory action by NRC in the future to provide for their availability, access to those records will depend on the voluntary cooperation of licensees.

Occupational Safety and Health Administration (OSHA) and National Institute for Occupational Safety and Health (NIOSH).--Access to employee health records is possible through voluntary cooperation of the employers, the issuance of a subpoena by OSHA or by NIOSH, or an OSHA or NIOSH request in accordance with OSHA regulations. Even when a subpoena has been issued by NIOSH, some employers have refused to supply requested employee records, arguing that employee privacy rights and the physician-patient testimonial privilege deny NIOSH the right to access without employee consent. The court decisions to date are split on whether NIOSH has the right to employer records for epidemiologic research without obtaining consent of the employee.

Employers using radation who are not regulated by NRC or an Agreement State* are subject to any health and safety standards that may be issued by OSHA. Current OSHA standards on ionizing radiation exposure and recordkeeping are similar to DOE and NRC requirements.

---

*See Glossary for definitions of technical terms.

9. Disease Registries

Although disease registries are a source of valuable data, there are three significant problems that limit their utility:

- Some holders of registries request recipients to guarantee that the information will not be redisclosed, and where Federal agencies are involved this cannot always be guaranteed because of the Freedom of Information Act;
- Some State privacy laws restrict the disclosure of information by registries operated by State agencies such as health departments;
- The VA has determined that, because of its confidentiality statute, it may not disclose names and addresses to registries not operated by certain public health authorities, thereby precluding VA participation in these registries. Some registries operated under contract from the National Cancer Institute (NCI) have been precluded by this statutory provision from receiving name and address information from the VA. The VA and NCI are presently trying to determine whether there is a basis for resolving the problem consistent with the VA statute.

10. Military Records

Military records are subject to the Privacy Act and are therefore available for research without restriction within DOD and for routine use by other researchers. No generally applicable routine use for epidemiologic research now exists.

11. Other Medical Records

Medical records maintained by private practitioners, hospitals, union health plans, employees, and insurance companies can assist epidemiologists in determining mortality or morbidity and in identifying relevant study groups. Access to these records depends on the voluntary cooperation of the record holders who are increasingly denying researcher access without the patient's consent or, if the patient is deceased, authorization from the next-of-kin.

GLOSSARY

1. Agreement State refers to the status of 25 of the 50 States. NRC has a program for transfer of some of its regulatory authority to the States. Under section 274 of the Atomic Energy Act, NRC may relinquish to the States its authority over the use of byproduct, source and special nuclear materials. NRC is required, however, to retain regulatory authority over the licensing of nuclear facilities such as reactors, exports and imports of nuclear materials and facilities, and larger quantities of fissionable materials. The mechanism for transferring NRC's regulatory authority is agreement between the Governor of a State and the Commission to do so. Before entering into an agreement, the Commission is required to make a finding that the State's radiation control program is compatible with the Commission's and that the State's program is adequate to protect the public health and safety. Thus far, 25 Agreement States have taken over the regulatory authority described above.

2. Bias is a type of error that causes a statistical estimate to deviate in a predictable direction from the true value.

3. Case-control studies are those in which an evaluation is made between two or more groups of the past differential exposure to an agent. The individuals of one group are selected on the basis of the presence of a specific disease or injury and are compared with individuals another group of "controls" selected on the basis

of the absence of this factor. An evaluation of the effect of previous differential exposure is attempted.

4. Cohort studies are those in which an evaluation is made of the differential incidence of disease within two or more groups. The individuals admitted to the studies are classified by level of exposure to a specific agent, with each group being followed over some period of time.

5. Cross-sectional studies are those in which an evaluation is made of the differential prevalence of disease within two or more groups. The individuals admitted to the studies are differentiated by amount or type of exposure for a specified period of time.

6. Ecologic studies are those in which some environmental factor is related to spatial and/or other patterns or morbidity or mortality in human populations. Health status is evaluated on the basis of aggregates of individuals as distinct from single individuals. An example is the comparison of cancer mortality in countries classified according to parameters such as the density of heavy industries, the average hardness of water, or background radiation.

7. Epidemiologic studies are those concerned with the distribution and determinants of disease and injury in humans, focusing on risks of disease in groups of individuals rather than on the presence or absence of disease in any single individual.

8. Rounding is the process of approximating a number by omitting certain of the end digits, replacing by zeros if necessary, and adjusting the last digit retained so that the resulting approximation is as near as possible to the original number. If the last digit is increased by unity, the number is said to be rounded up; if decreased by unity, it is rounded down. When both are under consideration, the process is said to be one of rounding off. In a computed result, the difference between that in which rounded numbers are used in the computations and that in which the precise representation of the number is used is called rounding error.

9. Sensitivity is defined as the percent of those subjects who actually have a given condition and who are detected as having such by a specific test procedure.

10. Specificity is defined as the percent of those subjects who are truly without the given conditions or who are "normal" and who are so classified by a specific test procedure.

*U.S. GOVERNMENT PRINTING OFFICE: 1980-0-722-821/499